

# AYP Consequences and Erasure Behavior

---

Vincent Primoli

5/7/2012

## Introduction

This paper asks the question, do the penalties facing schools and teachers affect the likelihood of irregularities. It examines the relationship between erasure behavior and Adequate Yearly Progress (AYP) status of schools. As consequences can dictate actions, it is presumed that aberrant erasure rates stemming from acts of impropriety can be influenced by the formalized system of accountability that is AYP.

The AYP statuses of schools in this study's data-frame are bucketed into five, general categories:

- **Made AYP (Made)** – school met all AYP criteria for all relevant subjects.
- **Level 1 (L1)** – school did not meet AYP criteria for first time in current year.
- **Level 2 (L2)** – school did not meet AYP criteria for at least two years. *Internal* changes are enacted to address problems.
- **Level 3 (L3)** – school did not meet AYP criteria for at least four years. *External* sources of assistance may be enacted to address problems.
- **Making Progress (MP)** – school met AYP criteria for first year of two-year probationary period and must repeat next year to be considered “back on track.”

These categories constitute a progression when examined across years; hence the directional relationships are also considered. For instance, roughly two-thirds of schools meet AYP standards in both of any two subsequent years in this study; because of this proficiency, one would expect this norm group to have lower incidence of aberrant erasure rates due to acts of impropriety. Schools showing directional AYP gains to compare with this norm group include MP-Made, L3-MP, L2-MP, and L1-Made. Particularly, in this study MP-Made (schools that move from Making Progress in one year to Made AYP the next year) and L3-MP (schools that move from Level 3 in one year to Making Progress the next) are found in starkest contrast to the majority that Made AYP in both years. Summary statistics and basic graphics are used to illustrate the relationships.

## Data Frame and Unit of Analysis - SGS

The data for this research comes from one large-scale assessment program over a two-year period across two subjects and seven grades. The unit of analysis is a school-grade-subject (SGS; e.g., School XXX, grade, 6, subject Math) as it is the smallest available unit reported. AYP reporting is at the school level, so every SGS within a given school will have the same AYP classification for a given year. Table 1 illustrates this for Example Junior High.

**Table 1: SGS AYP Data Example**

Example Jr. High						
SGS	School	Year	Grade	Subject	Erasures/Test	AYP Classification
1	1	2010	6	Math	0.8	Made AYP
2	1	2010	6	Reading	0.6	Made AYP
3	1	2010	7	Math	1.4	Made AYP
4	1	2010	7	Reading	0.6	Made AYP
5	1	2010	8	Math	0.7	Made AYP
6	1	2010	8	Reading	1.0	Made AYP

For each SGS, the data used in this study and explained later in text include the following:

- i) Erasure data
  - (1) SGS rates by erasure type (wrong-to-right, right-to-wrong, wrong-to-wrong) and test type (operational, field test)
  - (2) SGS outlier scores by erasure type and test type
- ii) AYP data
  - (1) School-level AYP classifications for previous eight years
- iii) Performance data
  - (1) SGS Z-scores – performance relative to program, grade-subject mean
- iv) Demographic data
  - (1) School-level percent of students eligible for free or reduced lunch (ECO %)

SGSs are grouped into four categories with the eight-year, historical AYP data. The first category, Made AYP, includes SGSs that have met AYP standards for all eight years, or the relevant duration for newer schools. The next category, Previous Level 1, implies at least one Level 1 classification over the previous, relevant years but no Level 2 or 3 classifications. Table 2 shows the percentage of SGSs by historical AYP classification.

**Table 2: SGS Historical AYP Classification**

AYP History	Percent of SGSs
Made AYP	42%
Previous Level 1	30%
Previous Level 2	14%
Previous Level 3	14%

So, 58% of SGSs failed to meet AYP requirements at some point over the last eight years and 28% failed to meet requirements for at least two consecutive years during that period. Those schools that failed for at least two consecutive years were subject to internal and/or external changes. In extreme cases these changes included staffing, so pressures and incentives were created to quickly bolster performance. Hence, it is reasonable to ask if aberrant erasure rates stemming from acts of impropriety are more likely for the SGSs with spottier AYP performance histories.

## WR Outlier Score

An average wrong-to-right (WR) erasure rate per test was computed for each SGS for comparison with a statewide baseline via Student's t-test. Though the untrimmed, erasure-rate data is not normal in the strictest sense (e.g. by Shapiro-Wilk), the generally larger SGS samples approximate the normal. Also, the T offers a straightforward way of rank-ordering the rates with consideration to sample variance and size. An SGS's p-value (p) given by the t-test is converted to an outlier score (OS) by the following log odds transformation:

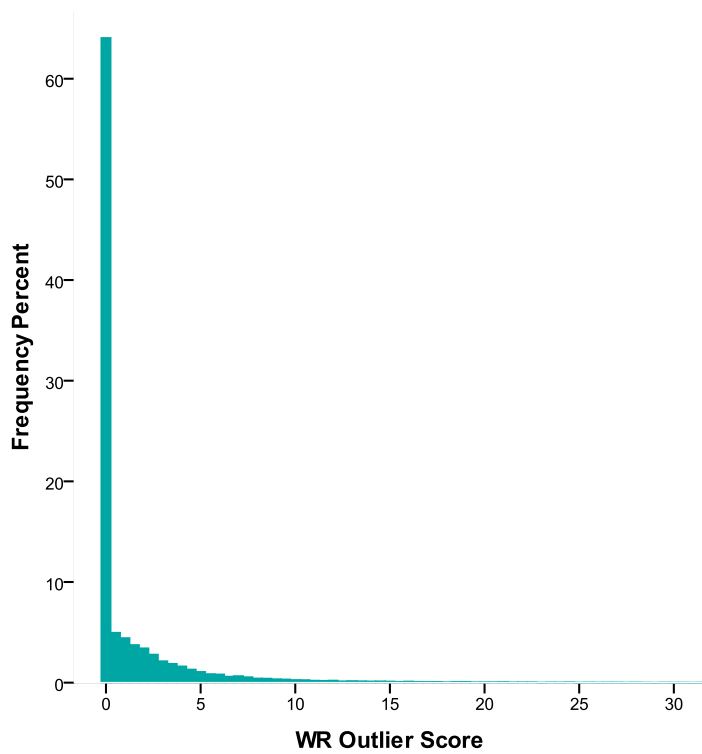
$$OS = \left| 1.086 \ln \left( \frac{p}{q} \right) \right|,$$

SGS's with a WR erasure rate less than or equal to the statewide baseline are given an OS of zero because the only concern is with identifying those with unusually large rates. Though somewhat arbitrary, the smallest OS threshold selected for identifying aberrances in this study is ten which is equivalent to a likelihood of .0001 under the T distribution. Table 3 shows the empirical percentage of SGSs above selected OS thresholds over two years for this particular large-scale assessment.

**Table 3: SGS Outlier Score Breakdown**

WR Outlier Score	% of SGSs
< 10	96.62%
>= 10	3.38%
>= 20	0.74%
>= 30	0.25%
>= 40	0.10%

**Figure 1: Histogram of SGS Outlier Scores**



## CWR Proportion

Total erasures (TE) are the sum of the following three kinds of erasures: wrong-to-right (WR), right-to-wrong (RW), and wrong-to-wrong (WW).

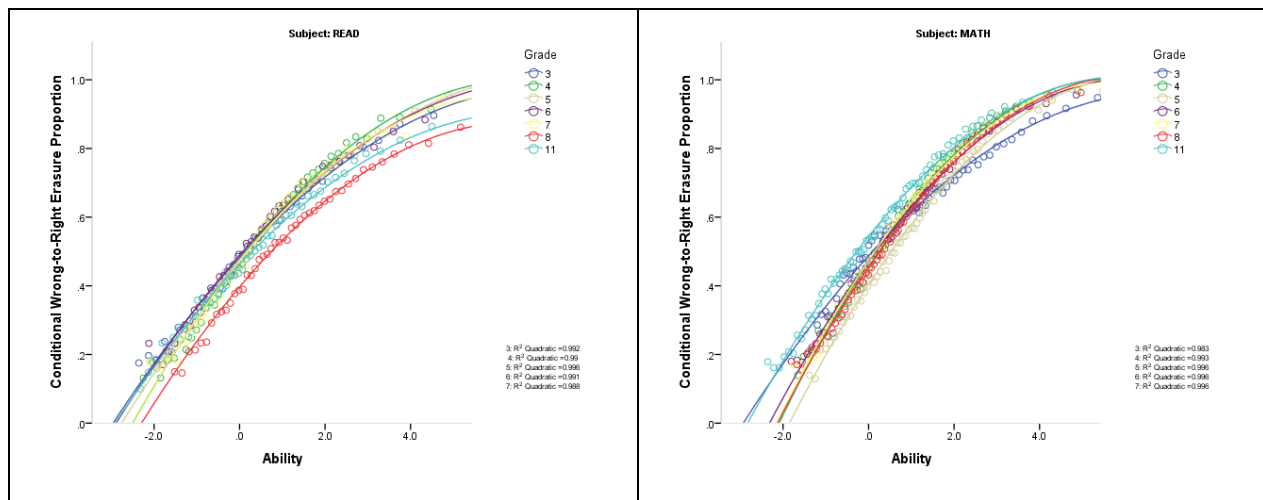
$$TE = WR + RW + WW$$

The conditional wrong-to-right erasure proportion (CWR) is the proportion of total erasures that result in a correct answer, or proportion of wrong-to-right erasures.

$$CWR = WR/TE$$

Expectations can be made regarding the proportion of all erasures that are wrong-to-right. Figure 1 plots observed, program-wide CWR by ability in both subjects and all grades. As correct answers imply ability, it is no surprise that CWR increases with ability.

Figure 2: CWR by Ability



The average CWR proportion is approximately .6 at the SGS-level. The correlation at the SGS-level is strong and positive ( $r = .52$ ), so CWR expectation increases or decreases with the average ability level of the SGS.

## Analysis & Results

Measures of economic status (ECO %) and school-level achievement (Z-Score) were introduced to better understand the make-up of the four subgroups created by historical AYP classification. The SGS means vary logically by the classifications as seen in Table 4. Most metrics for determination of AYP classification are related to performance. So, it is necessary that the average SGS performance, as measured by Z-Score, decreases by severity of historical AYP failure (i.e. Level 2 indicates more severe AYP failure than Level 1, etc). The previous section illustrated the relationship between CWR and ability; one expects higher CWR proportions for higher performing SGSs, and this holds. Also, SGSs with more AYP failure have a higher mean percentage of students eligible for free or reduced lunch (ECO %). Lastly, higher mean outlier scores are associated with more AYP failure.

**Table 4: SGS Means by Historical AYP Classification**

AYP History	Z-Score	ECO %	WR OS	CWR
Made AYP	0.54	30	0.8	0.66
Previous Level 1	-0.37	57	2.0	0.60
Previous Level 2	-0.80	70	2.8	0.57
Previous Level 3	-1.18	81	3.6	0.55

Conditional probabilities offer an answer to the question, does the likelihood of aberrant erasure rates vary by AYP historical classification. Table 5 clearly shows that the probability of aberrant erasure rates increases as AYP failure severity increases. Furthermore, the differences between Previous Level 3 and the reference group (Made AYP) are more pronounced for less probable erasure rates. For instance, an SGS with Level 3 classification was 12 times more likely to have an OS  $\geq 10$  than an SGS with a clean AYP history; the likelihood multiple associated with these two groups increased to 22 when considering OS  $\geq 30$ .

**Table 5: SGS Conditional Probabilities of Improbable WR Outlier Scores by Historical AYP**

AYP History	Percent of SGSs	>= 10	>= 20	>= 30	>= 40
Made AYP	42%	0.87%	0.18%	0.06%	0.00%
Previous Level 1	30%	2.74%	0.43%	0.08%	0.01%
Previous Level 2	14%	5.14%	0.83%	0.18%	0.04%
Previous Level 3	14%	10.39%	3.00%	1.24%	0.68%
Likelihood Multiple (Compared to reference group - Made AYP)					
		>= 10	>= 20	>= 30	>= 40
Previous Level 1		3.2	2.4	1.4	
Previous Level 2		5.9	4.6	3.3	
Previous Level 3		12.0	16.6	22.3	

Table 6 contrasts the erasure-rate differences in perhaps a simpler fashion. The columns total to 100%, and the five percentages in each row would be roughly equal if the groups were behaving similarly in terms of erasure rate. This is clearly not the case. SGSs with previous Level 3 violations are the most disproportionate; constituting only 14% of all SGSs, they account for 43% of all OSs >= 10, 56% of OSs >= 20, and so forth.

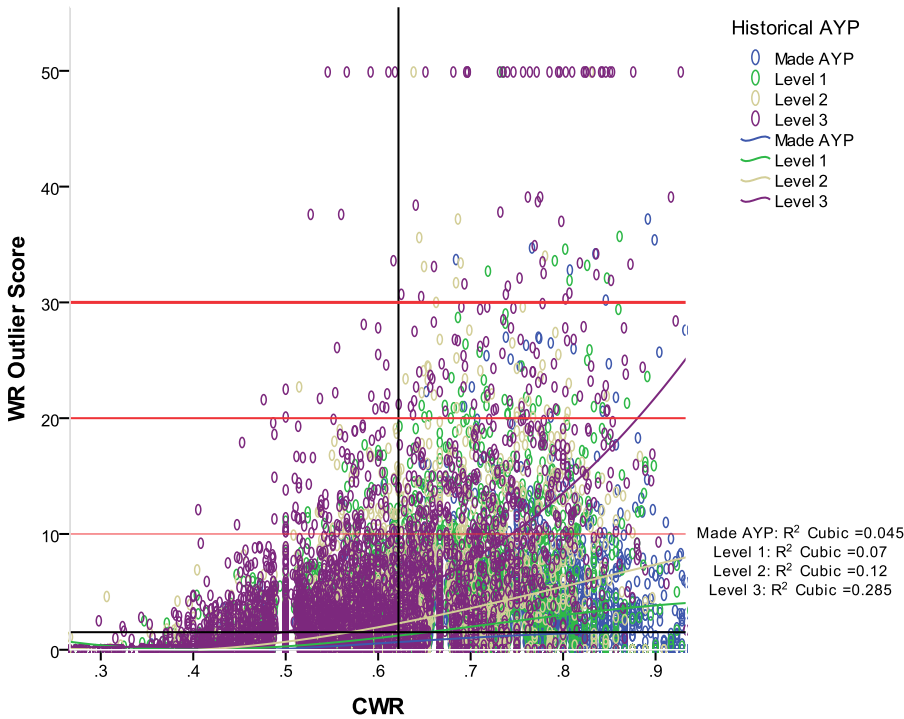
**Table 6: Percentage of Improbable WR Outlier Scores by Historical AYP**

AYP History	Percent of SGSs	>= 10	>= 20	>= 30	>= 40
Made AYP	42%	11%	10%	9%	0%
Previous Level 1	30%	24%	18%	9%	3%
Previous Level 2	14%	22%	16%	11%	6%
Previous Level 3	14%	43%	56%	71%	92%

In conducting data forensics that may have negative consequences for school systems and educators, it is advisable to employ multiple, statistical methods to determine which schools or classes to examine more closely. However, the plot in Figure 2 offers some good starting points on where one could begin investigating based upon erasures alone. The less likely SGS erasure rates, as measured by the OS, are higher on the y-axis, and SGSs with higher CWR proportions are to the right. Arguably, the most suspicious SGS is the top-right point which is around (.92, 50) and has a historical AYP classification of Level 3.



Figure 3: SGS Outliers Scores vs. CWR by Historical AYP



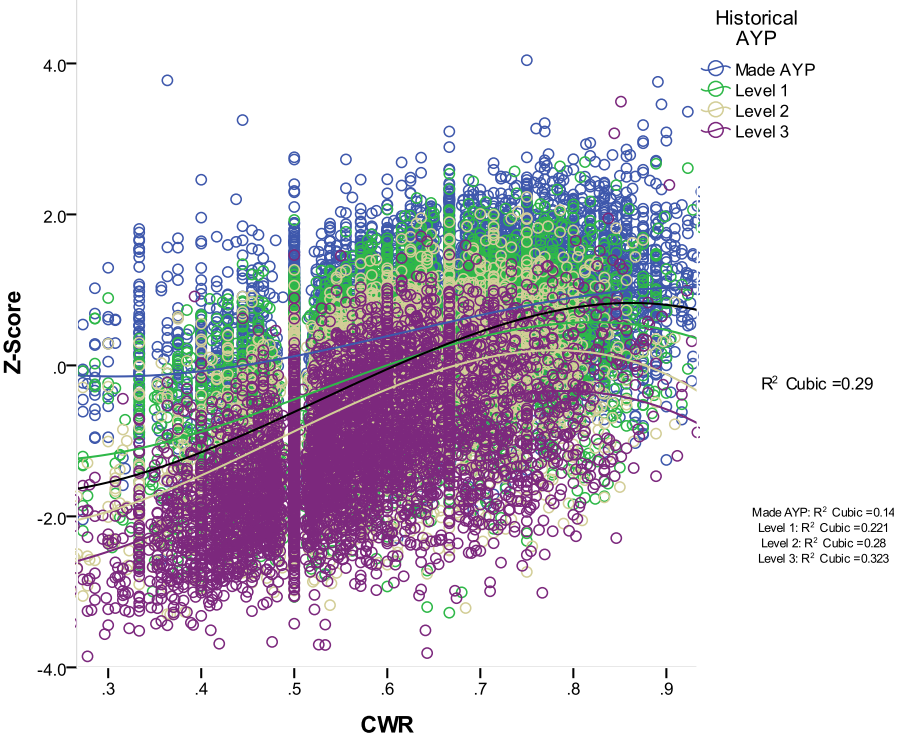
An interesting pattern emerges in Figure 2. A stronger correlation is observed between the two variables for SGSs with more historical AYP failure. The overall correlation is slight (about .2) but increases for all AYP failure levels; Level 3 exhibits a fairly strong correlation, around .53. So, Level 3 shows the strongest relationship between erasure proficiency (CWR) and likelihood of erasure rate. Cubic trends and the associated squared correlations are included by historical AYP group.

The centroid of these SGSs is at the intersection of black vertical and horizontal lines, the mean lines for each variable. The expectation for CWR for Level 3 schools, around .55, is substantially lower than the overall mean. So, most SGSs with OSs exceeding ten appear to erase more proficiently (high CWR) than is expected in their respective group. CWR is not stable for low erasure counts, but one could reasonably expect this proportion to converge with more erasures. Some very rare events such as multiple, extended misalignments on exams within an SGS could explain such proficiency, but very high CWR proportions raise suspicion when coupled with high erasure rates and lower performance.

Figure 3 illustrates the relationship observed between performance and erasure proficiency by AYP history. A stronger relationship is observed for SGSs with more AYP failure. The overall correlation

(about .54) is fairly strong, but ranges from .37 for SGSs with a clean AYP past to .57 for those in Level 3.

**Figure 4: SGS Performance vs. CWR by Historical AYP**



The previous analyses considered the entirety of each SGS’s AYP history for classification and comparison. But, AYP data can also be viewed from a directional perspective by combining two subsequent years of AYP categorizations. The two-year categorizations are shown in Table 7.

**Table 7: Two-Year AYP Categorizations**

AYP Losses	No Change	AYP Gains
Made-L1	Made-Made	MP-Made
L1-L2	L1-L1	L3-MP
L2-L3	L2-L2	L2-MP
MP-L2	L3-L3	L1-Made
MP-L3		

So, one can look for erasure behavior differences between classifications by using the erasure data associated with second year of the two-year classification. About two-thirds of SGSs classify as Made-Made in our data frame, indicating that they met AYP standards in both years. Table 8 shows how the SGS groups compare; all classifications contain more than 100 SGSs.

**Table 8: SGS Conditional Probabilities of Improbable WR Outlier Scores by Two-Year Directional AYP**

AYP Direction	Two-year Categorizations	Percent of SGSs	OS >= 10	OS >= 20	OS >= 30
Gain	MP-Made	3.3%	10.3%	3.5%	1.7%
Gain	L3-MP	2.0%	13.5%	2.4%	1.2%
Gain	L2-MP	2.7%	4.3%	0.9%	0.0%
Gain	L1-Made	4.9%	6.7%	1.7%	0.5%
No Change	L3-L3	4.5%	6.3%	1.9%	0.7%
No Change	L2-L2	1.3%	4.8%	0.0%	0.0%
No Change	L1-L1	0.4%	3.5%	2.1%	0.0%
<b>No Change</b>	<b>Made-Made</b>	<b>66.8%</b>	<b>2.0%</b>	<b>0.4%</b>	<b>0.1%</b>
Loss	MP-L3	1.6%	8.5%	2.2%	0.4%
Loss	MP-L2	1.0%	2.4%	0.0%	0.0%
Loss	Made-L1	8.7%	4.6%	1.0%	0.4%
Loss	L1-L2	1.9%	2.9%	0.5%	0.0%
Loss	L2-L3	0.8%	4.3%	0.0%	0.0%
	<b>Total</b>	<b>100.0%</b>	<b>3.4%</b>	<b>0.8%</b>	<b>0.3%</b>

The color scales apply column-wise and illustrate that higher incidence of erasure rate anomalies occur in SGSs making categorical AYP improvements; more red is seen in the top section of the table than in the bottom two sections. Schools have a two-year probationary period once they reach L2 or L3 AYP status. If they meet AYP standards, then they get moved to MP (Making Progress). To get back to a categorization of Made AYP, they need to meet the standards for two straight years. The AYP gains in Table 8 associated with MP have the highest incidence of aberrant erasure rates. So, it is important that such “progress” is examined with due scrutiny.

## Conclusion

Pressure to meet AYP standards could stimulate unethical behavior on the part of educators, and the numbers in this paper suggest an increased likelihood of aberrant erasure rates in probationary schools. In particular, schools with a long-history of AYP failure, hence subject to external sources of assistance including re-staffing, are most disproportionate in terms of extremely high erasure rates. AYP categorizations also revealed two interesting trends in correlation. First, schools with more AYP failure have stronger correlations between performance and erasure rate proficiency (CWR). Second, schools with more AYP failure have stronger correlations between erasure rate proficiency and erasure rate likelihood (OS). Finally, differences were observed when AYP categorization was made directional. Schools showing directional gains had much higher incidence of aberrant OSs.